Advanced Statistical Methods for Observational Studies

LECTURE 10

class management

- Bonus lecture:
  - We've always lost one lecture when we meet in person, but we were able to sneak in another lecture this year. Boom!
- We hope you've found this class useful.
- If you want to talk about any material or your own research feel free to reach out.

Mendelian randomization

AN INSTRUMENT?

Lawlor D, Harbord R, Sterne J, Timpson N, Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine 2008; 27:1133–1163.

Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Statistical Methods in Medical Research 2007; 16.

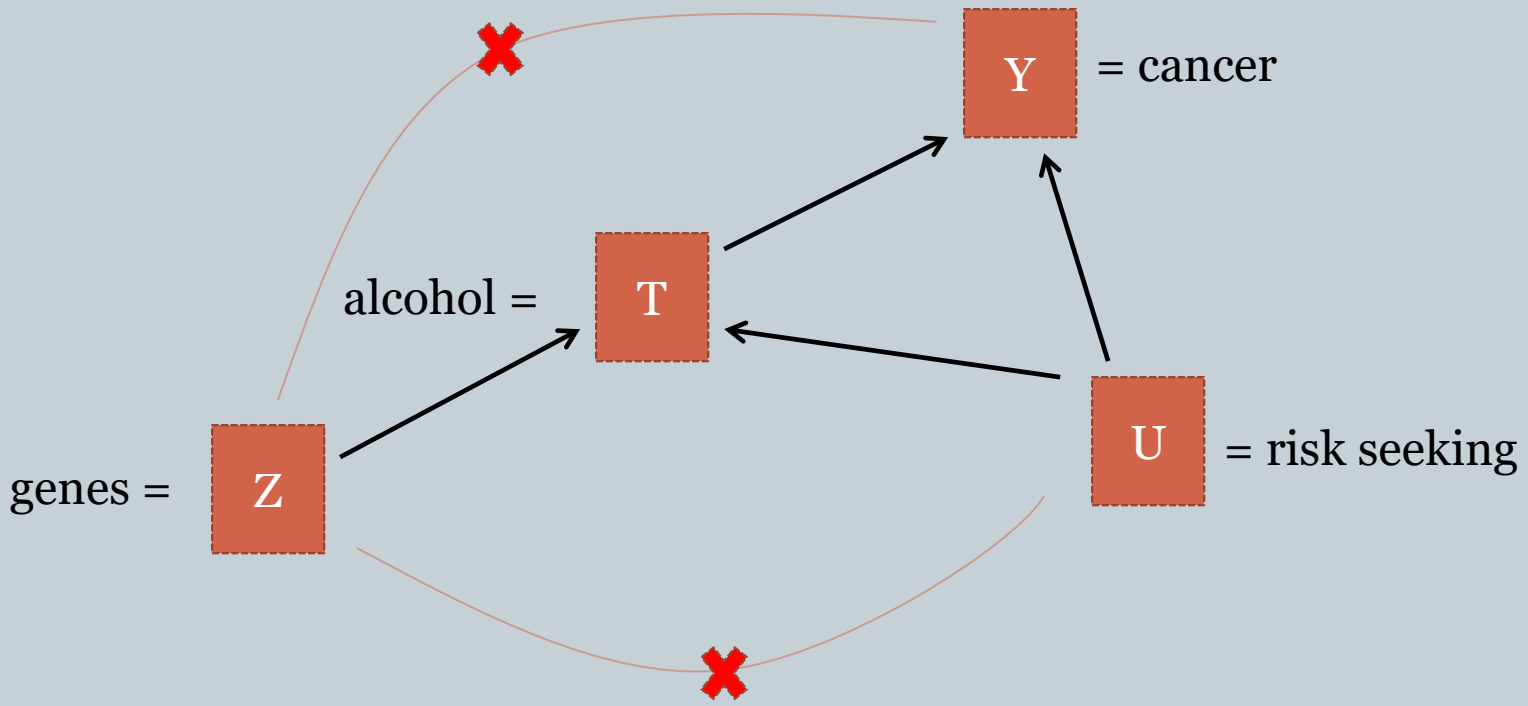
Mendelian randomization

- Mendelian randomization studies can be defined as any study that uses genetic assignment of a trait as a proxy for an environmentally modifiable exposure – used to make causal inferences about the outcomes of the modifiable exposure.

- Examples: C-reactive protein (CRP) and metabolic syndrome traits, alcohol consumption and oesophageal cancer, folate and schizophrenia, and folate and breast cancer.

Lawlor et al 2008

MR in a picture



Mendelian randomization

- These are usually epidemiological studies, often really helpful at addressing temporal ordering.
- Good as part of a portfolio of research on a causal effect, rarely viewed as definitive.
- Require solid understanding of gene-trait connections. You can imagine that old studies would need to be re-evaluated if the literature's understanding of gene-trait dynamics changes.

Lawlor et al 2008

Mendelian randomization: challenges

- Association with unmeasured confounders through **population stratification**:
  - Most Mendelian randomization analyses do not condition on parents' genes, creating the potential of the proposed genetic variant IV being associated with unmeasured confounders through population stratification.
  - If there are subpopulations, some of which are more likely to have the genetic variant, and some of which are more likely to have the outcome through mechanisms other than the treatment being studied.
  - Example: genes can be linked to social practices (e.g., diet).

Mendelian randomization: challenges

- Association with unmeasured confounders through **genetic linkage**:
  - Genes aren't totally independently assigned relative to other genes. If they are located near to each other on a chromosome then they are likely to be inherited together.
  - Example: Consider using a gene A as an IV. But, unfortunately, A is linked to a gene B which has a causal effect on the outcome through a pathway other than the treatment being studied. If gene B is not measured and controlled for, then gene A is not a valid IV because it is associated with the unmeasured confounder gene B.

Mendelian randomization: challenges

- Multiple pathways – "**pleiotropy**":
  - If the genetic variant being used as an IV affects the outcome through a function/pathway other than affecting the treatment being studied then this would mean the genetic variant has a direct effect on the outcome (violation of IV assumptions).
  - Example: want to study the causal effect of low-density lipoprotein cholesterol (LDLc) on myocardial infarction (MI) risk. The $\epsilon 2$ variant of the APOE gene is associated with lower levels of LDLc ("bad") but also associated with higher levels of HDLc ("good"). Thus, the gene APOE is pleiotropic, affecting myocardial infarction risk through different pathways, making it unsuitable as an IV to examine the causal effect of any one of these pathways on MI risk.

Mendelian randomization: takeaways

- Useful if reverse causality has been an issue in other epidemiological studies of this cause/effect relationship.
- Need to know the genetics really well, both the co-occurrences of other genes as well as the gene-to-trait connection(s).
- It is much stronger if you can use parent information (e.g., avoid population stratification issues). Ideally, sibling studies.

synthetic controls

LOL. WUT?

policy impact

- Example: After a major shooting in 1996, Australia implemented a number of policies to reduce firearm-related deaths.
- It seems sensible to ask: “what was the causal effect of these policies?”
- But, because this policy change impacts “everyone,” this is a challenging question to answer given the framework that we’ve discussed in this class.

Bartos et al 2020

policy impact: unit of observation

- In the regionalization examples (NICU and dissection), we had access to individuals' information and could think about individuals who were behaving as-if they were in a regionalized care setting, and contrast them with individuals who were not behaving as-if they were in a regionalized care setting.

- Again, here it's hard for two reasons:
  - SUTVA violations (interference)
  - We often don't have individual level data

policy impact

- Before the early 2000s, most researchers would deploy a case study to get at the causal effect of policies that operate on "large" units of aggregation (e.g., a city, a state, a country).
  - Select for regional variations.
  - Describe pre-policy settings.
  - Describe variation in policies (range: nothing, complete ban).
  - Describe details of implementations and politics.
  - Describe divergence in outcomes post.
- This. Is. Excellent. Research.

policy impact: synthetic control

- But can we quant that qual?
  - Yeah, that's what synthetic control designs were created to do.

- Basic idea: while it's hard (impossible?) to find a perfect counterfactual for a countrywide policy, maybe we can use several countries to get a sense of "on average" what a counterfactual would look like.

policy impact: synthetic control

Table 1 Non-zero weights for synthetic australia time series

Homicides		Motor vehicle fatalities		Suicide	
Nation	$w_i > 0$	*Nation*	$w_i > 0$	*Nation*	$w_i > 0$
Norway	0.305	Canada	0.597	Venezuela	0.393
United Kingdom	0.203	Venezuela	0.029	New Zealand	0.251
Singapore	0.178	Austria	0.374	Finland	0.136
New Zealand	0.141			Sweden	0.122
Costa Rica	0.097			Singapore	0.097
Canada	0.048				
USA	0.018				
Chile	0.010				
RMSPE	0.024	*RMSPE*	1.064	*RMSPE*	0.420

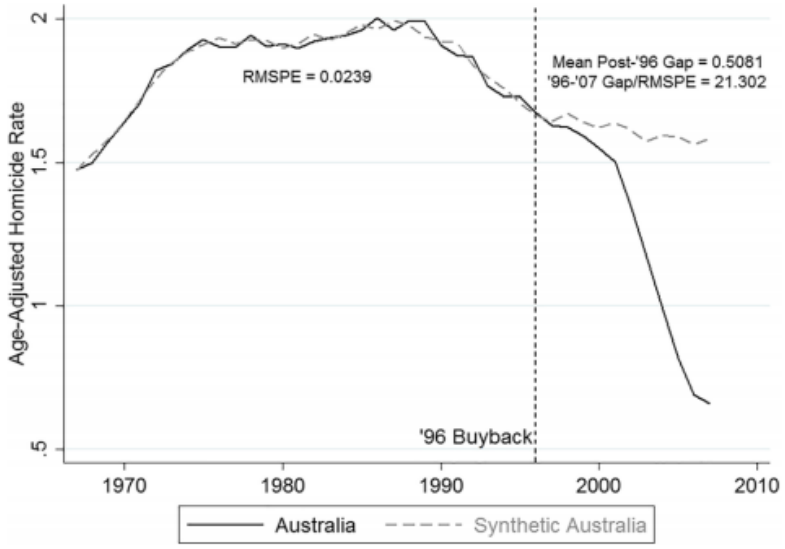


Fig. 1 Age-Adjusted Homicide Rates, Australia vs Synthetic Australia

Bartos et al 2020

synthetic control: issues

- How do we quantify the uncertainty?
  - Fundamentally, there's no source of variation we can locate
  - No one has come up with confidence intervals (maybe?)

- Two checks:
  - Time check: check a period where nothing should have happened and see if the procedure produces something spurious (this is a kind of "known null" check)
  - Control check: remove the "treated" and randomly choose one of the "controls" as your target and see what distribution of the test stat the method produces

Abadie, Diamond, and Hainmueller 2012

policy impact: synthetic control

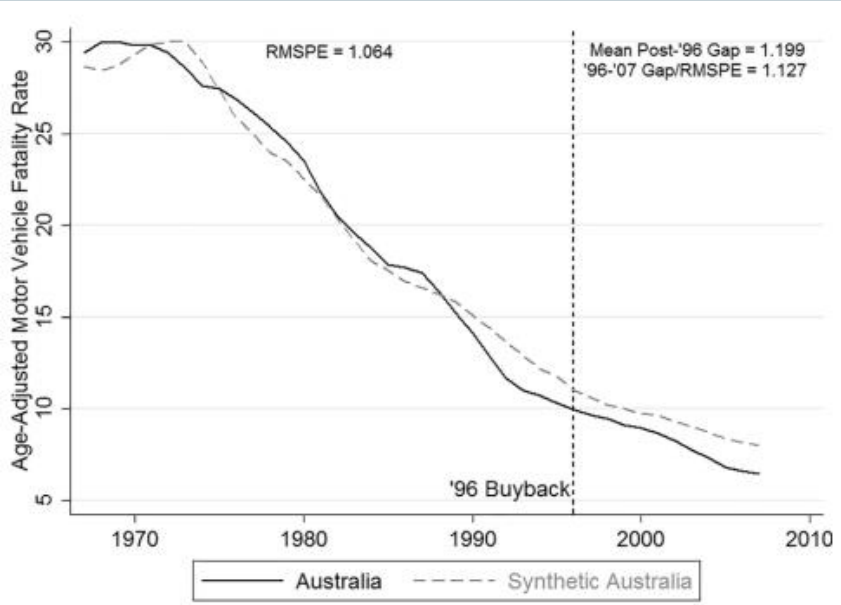


Fig. 2 Motor Vehicle Fatality Rates, Australia vs. Synthetic Australia

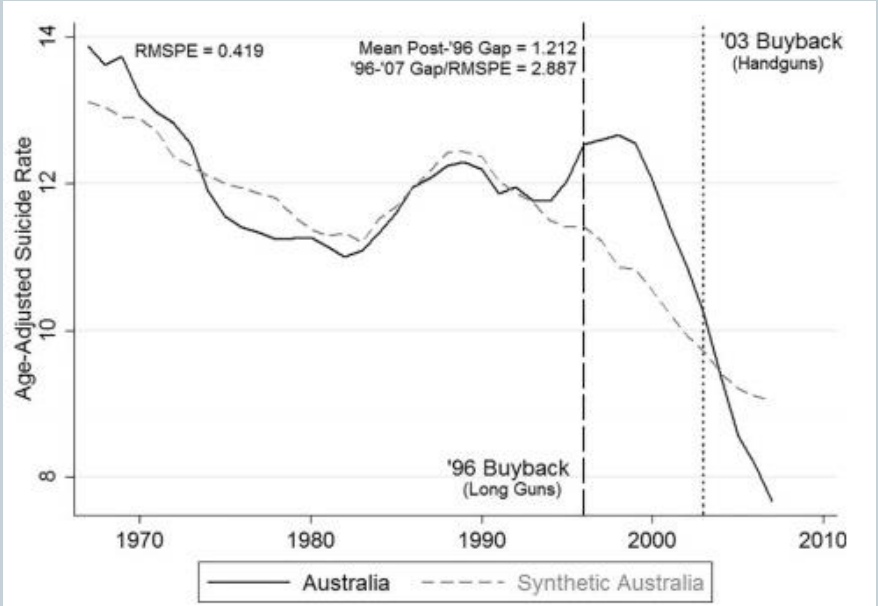


Fig. 3 Suicide Rates, Australia vs. Synthetic Australia

synthetic control: implementing

- Time divided into pre- (T_0) and post- (T_1) where $T_0 + T_1 = T$.

- Look at "outcomes" in the pre- period and c

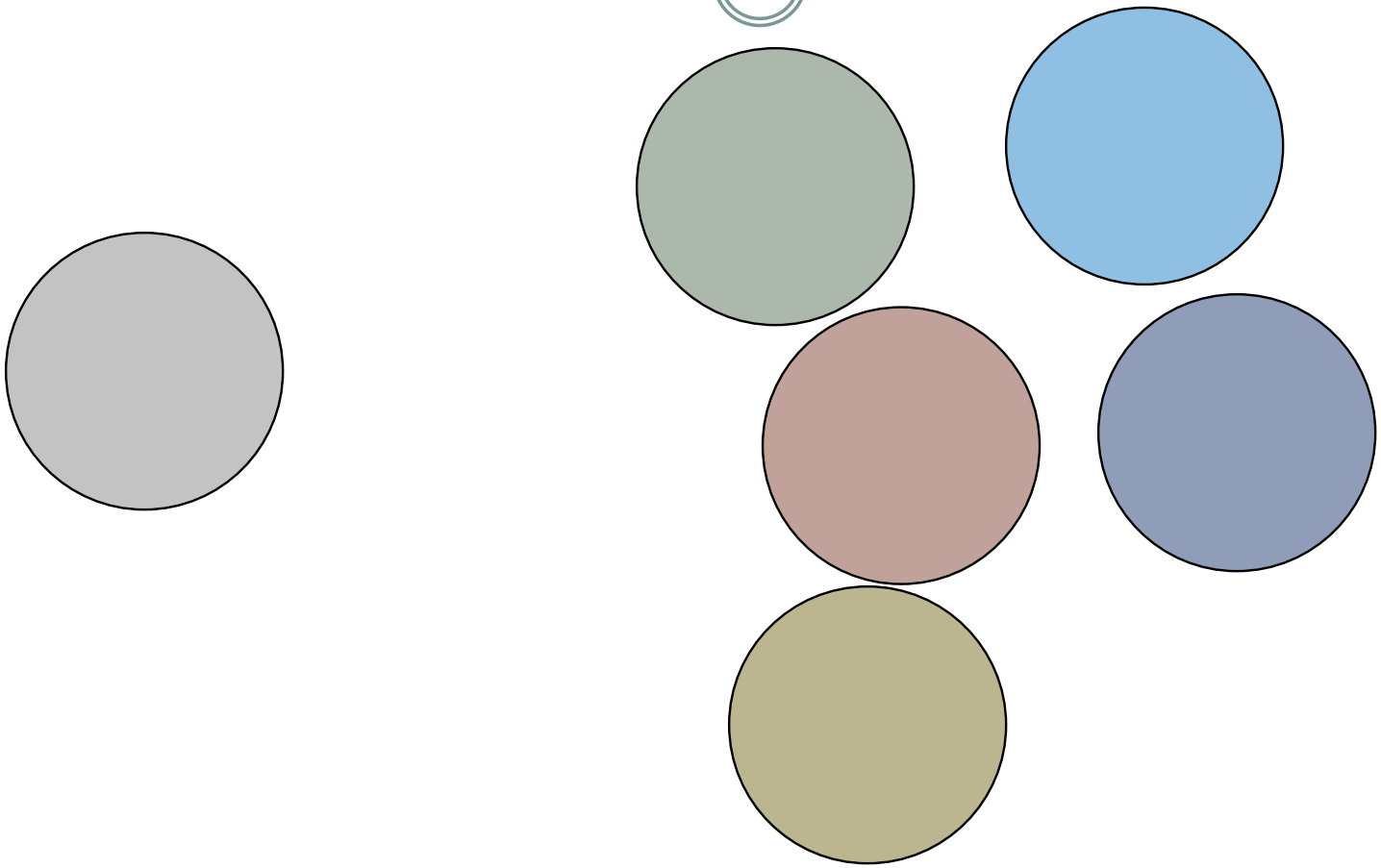
$$Y_{syn,t} = w_1 Y_{1t} + \cdots + w_J Y_{Jt}$$

where the w_j are selected by some criteria (usually..)

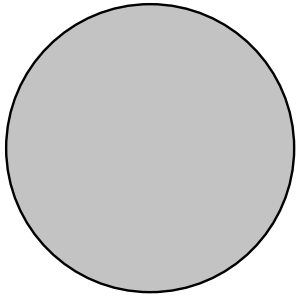
$$\sum_{t=1}^{T_0} \left(Y_{act,t} - Y_{syn,t}\right)^2$$

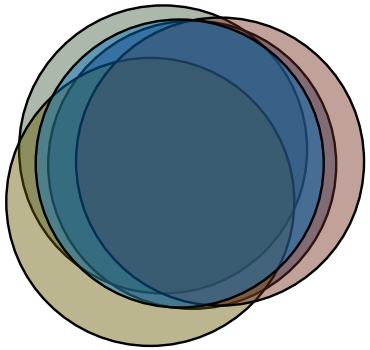
subject to $w_1 + \cdots + w_J = 1$.

Mike's silly visualization



Mike's silly visualization





synthetic control: Mike's issues

- What about interactions?
  - Say I have two chemicals – A and B – and I want to know what will happen if I mix them together and then apply heat.
  - Can I really understand what would happen to the mixture if I look at what happened when I applied heat to A and (separately) applied heat to B?
  - This is an issue with the "functional unit of observation."

- The consistency assumption (part of SUTVA):

$$Y_j(d,k) = Y_j(d)$$

where j is the unit, d is the exposure, and k is how the exposure was set to level d. Usually pretty simple. But gets messed up if multiple control-levels.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4912021/ (Hernan "Does Water Kill?")

synthetic control: useful and missing?

- A table clarifying the distribution of the controls' firearms policies:

homicides	weights	policies
Norway	0.305	A, B
United Kingdom	0.203	B, C, D
Singapore	0.178	A, B, C, D, E, F, G, H
New Zealand	0.141	B, C, D
Costa Rica	0.097	D, F, H
Canada	0.048	B, C, D, E
USA	0.018	lol.
Chile	0.010	E, F

fin.